

REMARKS

The Office Action dated June 3, 2009, has been received and carefully noted. The above amendments and the following remarks are submitted as a full and complete response thereto.

Claims 1-31 are currently pending and are respectfully submitted for consideration. By this Amendment, claims 1, 21, 24-26, 28, 29, and 31 have been amended and Claim 6 has been canceled, without prejudice and disclaimer. Support for the amendments to the claims can be found in the corresponding WO publication and the Specification, as filed. Thus, no new matter has been added.

In the Office Action, Claim 1 is objected to for informalities, as stated in point 6, page 4 of the Office Action. It is noted that Claim 1 has been amended and the amendments made to Claim 1 overcome the above objection. Thus, withdrawal of the objection of Claim 1 is respectfully requested.

Claims 1-6, 9, and 13-31 were rejected under 35 U.S.C. § 103(a) as being unpatentable over U.S. Pub. No. US 2004/0064447 to Simske et al. ("Simske") in view of U.S. Pub. No. US2004/0024760 to Toner et al. ("Toner"). Claims 7-8 were rejected under 35 U.S.C. 103(a) as being unpatentable over Simske in view of Toner, and further in view of U.S. Pub. No. 2002/0078024 to Bellany et al. ("Bellany"). Claims 10-12 were rejected under 35 U.S.C. 103(a) as being unpatentable over Simske in view of Toner and further in view of U.S. Pub. No. US2004/0181758 to Murakami et al. ("Murakami")

It is noted that Claims 1, 21, 24-26, 28, 29, and 31 have been amended. To the extent that the grounds for rejection are still applied to the currently pending claims, they are respectfully traversed.

In response to the arguments previously submitted, the Office Action asserts, on page 4, that different writing variations are taken into account in Simske. However, on page 7, the Office Action contradictorily asserts that Simske does not explicitly disclose that the data field is from a record and that the synonyms take into account writing variations. It is submitted that such contradiction is confusing and unclear.

Claim 1, as claimed, recites a method for tolerating writing variations in input data when processing a data record for finding a counterpart in a reference data set, the method comprising the steps of, among other things, "determining if a synonym candidate and the value of the data field fulfill a predetermined synonym acceptance criterion based on at least one quality parameter, wherein said at least one quality parameter takes into account writing variations that are evaluated based on differences in the value of the data field and the synonym candidate, and if the predetermined synonym acceptance criterion is fulfilled, associating the value of the data field and the synonym candidate as synonyms and automatically updating a synonym set representing known writing variations for the identifier in a computer readable database by adding the value of the data field to the synonym set without intervention of a user before searching for a counterpart, and searching for the counterpart for the data record by comparing the value of the data field to entries of the reference data set and/or the synonym set after the step of determining if the predetermined synonym acceptance criterion is fulfilled, wherein, if the synonym set was updated, said comparison to the synonym set comprises comparison to the updated synonym set in the computer readable database."

Similarly, Claim 21, as amended, recites a method of processing a synonym set stored in a computer readable database to tolerate writing variation in input data when searching counterparts in a reference data set for data records, a data record containing a data field representing an identifier, members of the synonym set being first identifier values and referring to respective second identifier values, the second identifier values being predetermined identifier values, and said searching for a counterpart involving comparison of a value of the data field to the synonym set in the computer readable database, the method comprising the steps of determining, by a processor, among the predetermined identifier values at least one synonym candidate relating to the value of the data field in the data record, and, if the value of the data field and a synonym candidate fulfill a predetermined synonym acceptance criterion based on at least one quality parameter, wherein said at least one quality parameter takes into account writing variations that are evaluated based on differences in the value of the data field and the synonym candidate, automatically adding before searching for a counterpart for a data record from the synonym set the value of the data field to the synonym set in the computer readable database as a member referring to the synonym candidate without intervention of a user.

Similar amendments are also included in other amended independent Claims 24-26, 28-29, and 31.

To better distinguish between the fields of the claimed invention and Simske, Claims 1, 21, 24-26, 28-29, and 31 are amended to define that the invention relates to tolerating writing variations in input data when processing a data record for finding a counterpart in a reference data set. Such amendment is supported, for example, on

page 1 of the WO publication, see in particular lines 5 to 9 and 24 to 32 and/or page 21, lines 19 and 20. Further, the step of determining if a synonym candidate and the value of the data field fulfill a predetermined synonym acceptance criterion is "based on at least one quality parameter, wherein said at least one quality parameter takes into account writing variations that are evaluated based on differences in the value of the data field and the synonym candidate." Examples of this kind of operation can be found explained in detail e.g. starting from page 16, line 4 to page 17, line 30.

Next, the step relating to the update is clarified by defining that the synonym set represents known writing variations for the identifier, as mentioned e.g. on page 3, lines 10 and 11.

This step is further clarified by correction where in the feature of adding the term "adding synonym candidate" is replaced by "adding the value of the data field" to bring the claims into consistency with the description, see e.g. page 6 lines 28-30 and page 10, lines 30, 31.

Further, the last step of amended Claim 1 and the same feature of other claims contained another passage are also amended to make it clear that it is the value of the data field that is compared to entries of the reference data set and/or the synonym set, as explained e.g. on page 3, lines 8 - 17 and page 9, lines 1 - 12 of the WO publication.

Applicants submit that the above amendments and the indication of support in the specification are intended to facilitate the Examiner to understand the invention and do not intend to narrow the scope of the claimed invention. No new matter has been added.

It is first noted that although both the claimed invention and Simske relate to searching, they relate to different fields of searching and are for use in different search applications. The invention, as is explained e.g. on page 1, see in particular lines 5 to 9 and 24 to 32, relates to a method and apparatus for tolerating writing variations in input data when processing a data record for finding a counterpart in a reference data set. An example of such operation is given on lines 20 to 22, where it is mentioned that new customer information (an example of the data record) may need to be checked against existing customer information (= an example of a reference data set) or against information obtained from official data registers (= another example of a reference data set). Instead of this, Simske relates to searching for desired information from a corpus of information, for example to searching by one or more search engines for a document from the Internet, see for example the abstract and paragraphs [0117] and [0118] on page 13 of Simske. Simsek does not give any indication or suggest anyhow that it could be used for tolerating/correcting errors in input data. The claims are now clarified in this regard.

According to the Examiner, Simske discloses generation and/or maintenance of a list of synonymous terms including the original input terms of the claims. However, Simske makes it clear that these synonymous query lists are for use in searching for documents, i.e. the generated lists provide a synonymous search query for a document search application. In this regards, a reference can be made, e.g. to paragraph [0081] on page 9, see in particular the sentence bridging the left and right columns, paragraph [00100] on page 11, paragraphs [0117] and [0118] on page 13 and Figure 7 of Simske.

In other words, in Simske the synonymic search query lists are used by at least one search engine that will search for all terms that are present in the list from a corpus or a great number of documents stored in various locations. The corpus or the documents, however, are not anyhow updated or stored anywhere in updated form as the result of the Simske operation. It is thus important to note that whereas in Simske a synonymic search query consisting of the base word and synonyms thereof is generated and then input into one or more search engines for use in searching e.g. from the Internet, in the claimed invention the used search term is the value of the data field (e.g. a misspelled street name from an input data record consisting a name and an address) and that in the claimed invention, when synonym candidates are identified, the search term as such is not anyhow changed or updated. Instead, the searchable synonym set is updated by the value of the data field (e.g. the misspelled street name) if the acceptance criteria is satisfied, where after a counterpart for the value of the data field is searched by comparing it to the updated searchable synonym set. Simske gives no suggestion or hint that the data to be searched could anyhow be affected by the generation of the synonymic search query and/or the subsequent search using said query. This is an essential distinction between the operating principles between the invention and Simske.

The amendment of the last feature of claim 1 should make the distinction between the current application and Simske, i.e. that the used incorrectly spelled input term can be compared against an updated set of searchable synonym data including said "acceptably" incorrectly spelled term that is stored and continuously automatically updated in a database. Instead of this, Simske simply generates a synonymic search

query based on an input term for use in searching, the synonymic search query consisting of a list of possible synonyms but no update (automatic or otherwise) of searchable databases takes place, and the search terms are not compared against any such updated databases.

Thus it is clear that, when considering the claims as whole, Simske does not disclose or suggest at least the combination of features that “if the predetermined synonym acceptance criterion is fulfilled, the value of the data field and the synonym candidate are associated as synonyms” and that thereafter “a synonym set representing known writing variations for the identifier are automatically updated in a computer readable database by adding the value of the data field to the synonym set without intervention of a user before searching for a counterpart by comparing the value of the data field to entries of the reference data set and/or the synonym set,” as recited in amended Claim 1, and similarly in amended Claims 21, 24-26, 28-29, and 31.

Furthermore, in view of the different writing variations and the quality parameter of the current independent claims, it is noted that, as is carefully explained in Simske, the weighting parameter thereof describes the “closeness in meaning”. In the particular example by Simske, the parameter thereof describes how frequently a particular synonym co-occurs with a given word, see for example paragraphs [0103] on page 12 and [0113] on page 13 of Simske. Paragraph [0103] of Simske explains that the synonyms are rated for their closeness in meaning or proximity to the original word. The proximity of Simske does not mean proximity in writing, but instead, as is clearly explained on page 13, paragraph [0113], lines 5 - 32, the proximity weighting thereof

refers to synonyms that are used in a document to define the same meaning without any consideration given how the words are written.

Thus, Simske does not disclose or suggest that any differences in writing should be considered, but only that similarities in meaning shall be analyzed. Therefore, there cannot be any disclosure or suggestion that Simske would disclose or suggest determination if a synonym candidate and a value of a data field fulfill a predetermined synonym acceptance criterion based on at least one quality parameter, the at least one quality parameter taking into account writing variations and being evaluated for the synonym candidate based on differences in the value of the data field and the synonym candidate, as recited in amended Claim 1, and similarly in amended Claims 21, 24-26, 28-29, and 31. Some examples for this type of operation are explained in detail e.g. from page 16, line 4 to page 17, line 30 of the WO publication of the present application.

Moreover, the present invention provides various advantages in searching for counterparts while tolerating writing variations in the input data. The data processing capacity for the actual searching is less than in prior art solutions where expanded lists of synonyms are compared to the searched data, as only the determined value of a data field with no expansion of the value is needed in the current invention. As a misspelled input term can be immediately accepted and the synonym set is automatically updated before looking for a counterpart for the term, the process is quick and yet tolerates misspelled or otherwise differently written terms as long as these satisfy the acceptance criterion. Since the method of the claimed invention does not consider the meaning, it is not dependent on the language used for the input but instead considers variations in the writing of the terms. Thus, similar algorithm can be used for any language. Because of

the lack of need to considerer meanings of the search terms, which requires considerable computing capacity, the requirement for computing capacity is less than e.g. in Simske where such considerations are crucial. The database of searchable possible writing variations to which values of the data fields of subsequent searches are compared to grows automatically without any user intervention, and thus in time the database, for example address register, grows to a considerable size and thus different writing variations in the input data can be tolerated also for this reason.

As an example of thus, Applicants enclose an excel sheet (three pages) where a real life example relating to a name and address register is shown. Page 1 shows how a searchable database of 571,351 terms at the beginning (year 2003) has grown to 1,458,684 terms by 25 August 2009. Although apart of the growth is due to addition to the official street name record (the reference data set), the majority is due to the automatic machine discovery base on input data. This part of the growth has been provided without human intervention by means of the present invention. The second page shows 42 street names in the register with most synonyms in the synonym set. The third page shows the history of the six top most terms since 2003 to date. As shown, the number of acceptable synonyms has grown automatically each year based on customer inputs without any intervention by the operators.

The above information clearly shows that the claimed invention has commercial success, which should be allowable over the cited art.

Further, as previously submitted, none of Toner, Bellany, and Murakami, taken alone or in combination, cures the deficiencies of Simske. Accordingly, it would not have

been obvious for one skilled in the art to combine Simske and Toner to achieve the subject matter of the claims.

Based on the above, Claims 1, 21, 24-26, 28, 29, and 31, as amended, are allowable over the cited art.

At least based on the above reasons, Claims 2-5, 7-20, 22-23, 27, and 30, which depend from Claims 1, 21, 27, 28, or 29, are likewise allowable at least due to their dependencies from allowable independent claims and additional features recited therein.

Conclusion

In view of the foregoing, reconsideration of the application, withdrawal of the outstanding rejections, allowance of the currently pending claims, and the prompt issuance of a Notice of Allowability are respectfully solicited.

Should the Examiner believe anything further is desirable in order to place this application in better condition for allowance, the Examiner is requested to contact the undersigned at the telephone number listed below.

In the event this paper is not considered to be timely filed, Applicant respectfully petitions for an appropriate extension of time. Any fees for such an extension, together with any additional fees that may be due with respect to this paper, may be charged to counsel's Deposit Account No. 01-2300, **referencing docket number 108800-00007.**

Respectfully submitted,



Wan-Ching Montfort
Registration No. 56,127

Customer No. 004372

ARENT FOX LLP

1050 Connecticut Avenue, N.W., Suite 400

Washington, D.C. 20036-5339

Tel: (202) 857-6000

Fax: (202) 638-4810

GEO/CYM

Enclosure: Exemplary Information Sheets (3 pages).

Year	Description	number
	Level 0 - Synonyms added as human work starting from 1980's. People had synonym tools that shows candidates but decision was made by individual human. (As tele operator 2003 customer service tools even today are).	571,351
2003	Batch job level machine discovery based on 571351 existing synonyms.	416,298
2003	Growing different variations based on machine discovery model.	113,261
2003	Automated machine discovery based on individual / customer input data	3,729
2003	Automated machine discovery based on official street name growth	
2004	Automated machine discovery based on individual / customer input data	83,865
2004	Automated machine discovery based on official street name growth	5,462
2005	Automated machine discovery based on individual / customer input data	57,471
2005	Automated machine discovery based on official street name growth	5,425
2005	Manual +	206
2005	Manual +	1
2006	Automated machine discovery based on individual / customer input data	60,768
2006	Automated machine discovery based on official street name growth	5,877
2007	Automated machine discovery based on individual / customer input data	56,104
2007	Automated machine discovery based on official street name growth	5,316
2008	Automated machine discovery based on individual / customer input data	38,805
2008	Automated machine discovery based on official street name growth	5,560
2009	Automated machine discovery based on individual / customer input data	24,189
2009	Automated machine discovery based on official street name growth	4,996
		1,458,684
	25/08/2009	

Street name	Number of synonyms
Wilhelmschildtinkatu	333
Nordenskiöldinkatu	301
Adlercreutzinkatu	220
Munkkiniemenpuistotie	211
Aleksanterinkatu	208
Mannerheimintie	204
Runeberginkatu	198
Nordenskiöldintie	197
Opiskelijankatu	196
Kalevankatu	193
Cygnaeuksenkatu	180
Adolfliindforsintie	179
Lönnrotinkatu	179
Saarijärventie	177
Väinämöisenkatu	177
Pohjolankatu	177
Svanströminkuja	175
Stålarminkatu	174
Rantatie	174
Kauppakatu	171
Mechelininkatu	170
Svinhufvudintie	169
Merirastilantie	169
Jäkälänpuistokatu	164
Tapanilöfvinginkatu	164
Virontörmänskatu	164
Albertpetreliuksenkatu	163
Itäinenpitkäkatu	162
Schaumaninpuirostie	161
Hakalantie	159
Vanhasuutarinkyläntie	159
Itsenäisyydenkatu	159
Minnacanthinkatu	158
Merimiehenkatu	157
Karjalantie	157
Käsityöläiskatu	155
Lauklähteenkatu	155
Käsityöläisentie	153
Pohjoinenliipolankatu	152
Taitoniekantie	152
Kauppakartanonkatu	151
Karjalankatu	151

Katu	2003	2004	2005	2006	2007	2008	2009	Total
Wilhelmschildinkatu	194	53	23	19	19	16	9	333
Growth (yearly / total)	27%	9%	7%	7%	5%	5%	3%	72%
Nordensköldinkatu	180	32	21	26	24	16	2	301
Growth (yearly / total)	18%	10%	11%	9%	6%	1%	67%	
Adlercreutzinkatu	151	21	9	18	7	8	6	220
Growth (yearly / total)	14%	5%	10%	4%	4%	3%	3%	46%
Munkkiniemenpuistotie	154	15	12	9	10	9	2	211
Growth (yearly / total)	10%	7%	5%	5%	5%	5%	1%	37%
Aleksanterinkatu	123	24	15	14	17	9	6	208
Growth (yearly / total)	20%	10%	9%	10%	5%	5%	3%	69%
Mannerheimintie	143	10	14	10	13	10	4	204
Growth (yearly / total)	7%	9%	6%	7%	5%	5%	2%	43%